

RECEIVED
TECH CENTER 1600/2900
02 FEB 20 PM 3:41

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Deb K. Chatterjee

Appl. No.: 09/558,421

Filed: April 26, 2000

For: **Mutant DNA Polymerases and Uses
Thereof**

Art Unit: 1652

Examiner: Rao, M.

Atty Docket: 0942.3600003/RWE/BJD

#9
A.9.3
267/02

Declaration of Harini Shandilya

Commissioner for Patents
Washington, DC 20231

Sir:

I, Harini Shandilya, DO HEREBY DECLARE AND SAY:

THAT, I am presently employed by Life Technologies, Inc. (hereinafter "Life Technologies," and now Invitrogen Corporation)¹, the assignee of the above-captioned application as a Scientist. I have an M.S. degree in chemistry. I have been employed by Life Technologies since June 1, 1988.

THAT, one of my supervisors at Life Technologies has been Deb K. Chatterjee. During the time period the experiments described in this Declaration took place, I worked in close physical proximity to Deb Chatterjee. During this time period, I had at least a generalized knowledge of Deb Chatterjee's research projects.

THAT, at times Deb Chatterjee would ask me to witness his laboratory notebooks. When he asked me to witness laboratory notebook pages, I would carefully read and understand the content of those pages, then sign and date the pages at the lower lefthand corner. Deb

¹Life Technologies, Inc. merged with Invitrogen Corporation on September 12, 2000, with Invitrogen Corporation being the surviving entity.

RECEIVED
TECH CENTER 1600/2900
FEB 25 2002

Chatterjee's notebook would often contain original data, e.g., photographs of gels. When original data was included in notebook pages that I witnessed, I would review the data together with the remaining content of the pages.

THAT, prior to October 17, 1994, I was aware that Deb Chatterjee had worked on a research project relating to mutant DNA polymerases. I was aware of this research project from various sources. These sources included informal discussions in the laboratory with Deb Chatterjee, discussions I heard between Deb Chatterjee and others, and my review of Deb Chatterjee's notebook, including original data included in that notebook.

Attached to this Declaration are several Exhibits containing laboratory notebook pages recorded by Deb Chatterjee. During the time period these experiments were conducted, as indicated on the notebook pages, I saw Deb Chatterjee performing molecular biology experiments. The nature of these experiments is consistent with the notebook entries and the cloning of a mutant DNA polymerase gene.

Attached to this Declaration as Exhibit A is a copy of a laboratory notebook page recorded by Deb Chatterjee. At the bottom lefthand corner of this notebook page I recognize my signature, indicating that I witnessed and understood this page on the date indicated. While the date is masked on the attached Exhibit, the date on the original is before October 17, 1994. Therefore, I conclude that this notebook page was in existence, and its content was communicated to me, before October 17, 1994.

The content of Exhibit A relates to a research plan to make *Taq* DNA polymerase mutants. I note that this page indicates that a distinctive difference between T7 polymerase, which utilizes dNTP and ddNTP equally well, and *Taq* polymerase in the presence of a Tyr in

lieu of a Phe at a location in the O-helix corresponding to Phe₇₆₂ of *E. coli* polymerase I (Klenow fragment). This page also indicates to me that *Taq* DNA polymerase mutants would be made based on these differences, in order to construct a *Taq* polymerase that is more T7-like, e.g., would use dNTPs/ddNTPs equally well. I conclude that this page teaches a plan for making a *Taq* DNA polymerase mutant containing Tyr in lieu of Phe at a position corresponding to the Phe₇₆₂ of *E. coli* polymerase I (Klenow fragment). Thus, the concept of this mutation had been communicated to me prior to October 17, 1994.

Attached to this Declaration as Exhibit B is a copy of two laboratory notebook pages recorded by Deb Chatterjee. At the bottom lefthand corner of the notebook page I recognize my signature, indicating that I witnessed and understood the page on the date indicated. While the date is masked on the attached Exhibit, the date on the original beside my signature is before October 17, 1994. Therefore, I conclude that these notebook pages were in existence, and their content was communicated to me before October 17, 1994.

The content of Exhibit B relates, among other things, to *Taq* mutant F667Y. The nomenclature "F667Y" refers to a substitution of Tyr (Y) for Phe (F) at position 667 of the *Taq* polymerase. This substitution is the same substitution referred to in Exhibit A. However, in this Exhibit the exact amino acid residue of *Taq* polymerase, instead of the corresponding *E. coli* polymerase I (Klenow fragment) residue, is explicitly identified. Thus, the concept of a specific mutation, Phe₆₆₇ → Tyr₆₆₇ of *Taq* polymerase, was conveyed to me prior to October 17, 1994.

Exhibit B also describes a mutagenesis experiment that was actually performed. In this experiment, oligonucleotide 2680 was annealed to a single-stranded DNA containing the wild-type *Taq* polymerase gene at 70°C for 2 minutes followed by cooling to about 35°C over

30-40 minutes. The resulting hybridized molecule was incubated with T7 polymerase and T4 ligase to synthesize and ligate the strand complementary to the single-stranded DNA using oligonucleotide 2680 as a primer, and then used to transform bacteria. DNA was then isolated from the transformed bacteria. This DNA was assayed for the presence of an additional *AseI* restriction site derived from the oligonucleotide sequence. The presence of this additional restriction site would indicate incorporation of the oligonucleotide sequence, containing the F667Y mutation, into the mutagenized clone. Page 2 of Exhibit B contains original data showing the results of an assay of various mutagenized clones for the presence of the additional 590 bp *AseI* restriction fragment, which would indicate the presence of the *AseI* restriction site derived from the oligonucleotide sequence.

This page of Exhibit B is signed by me as having been witnessed and understood by me prior to October 17, 1994. Therefore, I would have reviewed this original data prior to October 17, 1994. Thus, experiments designed to construct this mutation and test it for the incorporation of an additional *AseI* restriction site were performed prior to October 17, 1994.

Attached to this Declaration as Exhibit C is a page from Deb Chatterjee's notebook depicting an assay of additional clones for the presence of the additional *AseI* restriction site. I have independently reviewed the original data attached to this page of Deb Chatterjee's notebook. Based on this independent review, I conclude that clone 8 contains an additional *AseI* restriction site.

THAT, I have reviewed the oligonucleotide sequence identified as No.2680 on the Request for Synthesis of Oligonucleotides, which is attached to this Declaration as Exhibit D. I have also reviewed the publication of Lawyer et al., J Biol. Chem. 264(11):6427-6437 (1989),

which is attached to this Declaration as Exhibit E. In Figure 2 of Lawyer *et al.*, the DNA sequence of the *Taq* polymerase gene is disclosed. I have compared the DNA sequence in Figure 2 of the Lawyer article surrounding codon 667 with the sequence of oligonucleotide 2680 of Exhibit D to this Declaration. Based on this comparison, I have determined that the two sequences are complementary with two exceptions: (1) oligonucleotide 2680 contains a sequence complementary to a Tyr codon instead of the Phe codon in the sequence of Figure 2; and (2) oligonucleotide 2680 contains a sequence corresponding to an *AseI* restriction site not found in the sequence of Figure 2.

THAT, I hereby declare that all statements made herein of my own knowledge are true; and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code; and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

2/15/02
Date

H Harini Shandilya

No.

Annealing:

	Kinased Oligo 2680	Kinased oligo 2723	Control
Amn. Buffer	1μl	1μl	1μl
U-ssDNA (gene cleaved)	1.5μl	1μl	1.5μl
H ₂ O	6.5μl	6.5μl	7.5μl

Anneal at 70°C for 2min → Let cool to ~35°C over 30-40 min.

Synthesis:

Add to tubes:

1μl 10X Syn. Buffer
0.3μl T7 DNA Pol (usB)
1.0μl T4 ligase (LT1)

5' on ice, 5' at room temp → 45 min at 37°C.

Add 90μl TE (10mM Tris pH8.0, 10mM EDTA)

Transfect DH5⁺ E. coli

Transform DH5⁺ E. coli 100μl with 1.5μl ligation/synthesis mix.

30 min on ice

42°C for 40 Sec. → Split 10% and 90% for 2680 and 2723
Keep 100% for the control.

Add 60μl lawn cells + 3ml soft agar

→ Let the agar solidify

→ Incubate plates at 37°C

Assessed & Understood by me,

Date,

Invented by

Date

Recorded by
John Doe

To Page No.

On Page No. _____

of Plaques in the plates

2680 (F667Y)	10%	90%	100%
	285	TMTC	-

2723 (Y671S)	250	TMTC	-
--------------	-----	------	---

Control

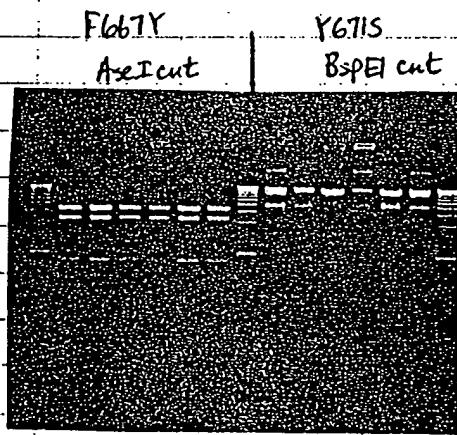
150

Test 5 plaques from each, 50µl lawn cells + Agar/Plaque Slab.
Incubate at 37°C for 6 hrs!

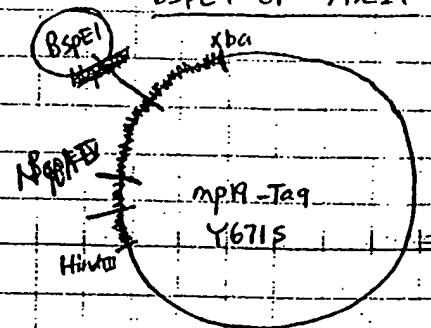
Standard mini prep - Take DNA in 10µl TE

1µl was (digested w) F667Y with AseI NEB Reut 3

1µl (") Y671S with BspEI NEB Reut 3.



* Control RF cut,
BspEI or AseI.



Y: AseI cut : expected ~590 bp fragment from one of the two larger fragment : Thus, are correct clone — Recheck few more clones.

S: BspEI cut : Expected a linear frag. # 4 is the correct clone. # 3 has an band (Save both).

To Page 1

Assessed & Understood by me,

Date

Invented by

Date

Recorded in, 11/11/01

10% plating > 500 colonies
90% plating in T.C.

Test 6 colonies

- * Control pTTR18-Tag parent cut with BspEI. - it does not have any BspEI.

#1, 3, 4, 5 & 6 are correct as they have the BspEI site (that was created). P

The plasmids were pooled.

Save #1 & #3 as

Grow #1 & #3 for enzymic activity.

67Y: Test 8 more plagues.

1 → 8

六

As e I
cut

* #8 seems to be mixture of wild type and the mutant. The mutant should produce two bands ~ 1450 bp and ~ 1850 bp from the second largest band.

Thus #8 was saved. The fragment will be cloned as NgoAI + Xba and cloned into pTQtag. Then, look for an extra AseI site.

To Page No.:

Wth a Kiss & Understood by me,

Date

Invented by

Date

11.1

Recorded by *Bob B. Goldsby*

REQUEST FOR SYNTHESIS OF OLIGONUCLEOTIDES

① 12649 ✓
② #2680 -

Dates

Names

Department:

Account 1

Microbiology

Costs

~~\$180~~

① GAC | GGG | GAT | GTT | CGC | GAG | GTT | GGG | ATC (#2649
24.mv)

② GTA | GAG | GAC | CCC | GTA | ATT | AAT | GGT | CTT | GGC | CGC
A_{ac}T 12680
(33. mar)

Brief Description Of Use.

#1 Q582A mutation of Tag

#2 F667Y mutation of Tag.

#2 F667Y mutation of Tag.

APPROVAL:

Dr DIRECTOR OR VICE PRESIDENT

F

DATE RECEIVED

三

DATE SYNTHESIS COMPLETED

Isolation, Characterization, and Expression in *Escherichia coli* of the DNA Polymerase Gene from *Thermus aquaticus**

(Received for publication, August 31, 1988)

Francis C. Lawyer†, Susanne Stoffel‡, Randall K. Saiki§, Kenneth Myambot, Robert Drummond¶, and David H. Gelfand†,||

From the Departments of †Microbial Genetics, §Human Genetics, and ¶Protein Chemistry, Research Division, Cetus Corporation, Emeryville, California 94608

The thermostable properties of the DNA polymerase activity from *Thermus aquaticus* (*Taq*) have contributed greatly to the yield, specificity, automation, and utility of the polymerase chain reaction method for amplifying DNA. We report the cloning and expression of *Taq* DNA polymerase in *Escherichia coli*. From a λ gt11:*Taq* library we identified a *Taq* DNA fragment encoding an epitope of *Taq* DNA polymerase via antibody probing. The fusion protein from the λ gt11:*Taq* candidate selected an antibody from an anti-*Taq* polymerase polyclonal antiserum which reacted with *Taq* polymerase on Western blots. We used the λ gt11 clone to identify *Taq* polymerase clones from a λ Ch35:*Taq* library.

The complete *Taq* DNA polymerase gene has 2499 base pairs. From the predicted 832-amino acid sequence of the *Taq* DNA polymerase gene, *Taq* DNA polymerase has significant similarity to *E. coli* DNA polymerase I. We subcloned and expressed appropriate portions of the insert from a λ Ch35 library candidate to yield thermostable, active, truncated, or full-length forms of the protein in *E. coli* under control of the *lac* promoter.

Taq Pol I only at the beginning of the PCR reaction rather than before each round of amplification.

A 62–63-kDa *Taq* Pol I has been purified from *T. aquaticus*, but growing the organism is more difficult than *E. coli* and polymerase yields are low (4, 5). We have developed an alternative purification protocol² yielding a 94-kDa enzyme with 10–20 times higher specific activity than that previously reported. While the activity yield is quite high (40–60%), the initial expression level of *Taq* DNA polymerase in the native host is quite low (0.01–0.02% of total protein). Therefore, we sought to clone the *Taq* Pol I gene and express the gene in *E. coli*. In addition, the availability of the enzyme and the DNA sequence of the *Taq* DNA polymerase gene will facilitate the study of structure/function relationships and permit detailed comparisons with mesophilic DNA polymerases.

MATERIALS AND METHODS³

RESULTS

λ gt11 Libraries—The construction of three λ gt11:*Taq* libraries is described under "Materials and Methods," in the Miniprint. To maximize the probability of recovering a *Taq* Pol I epitope, three separate *Au*I libraries were prepared. We ligated 8-mer, 10-mer, and 12-mer *Eco*RI linkers to the *Taq* *Au*I DNA fragments to ensure that each *Au*I fragment would be in-frame with respect to β -galactosidase in one of the libraries. Upon screening with primary antibody from *Taq* Pol I-immunized rabbits and plaque purification, we identified seven positive plaques from the 12-mer library, four positive plaques from the 10-mer library, and no positive plaques from the 8-mer library. The *Eco*RI inserts fell into four size classes: two of the seven phage isolated from the 12-mer library and two of the four phage isolated from the 10-mer library contained 115-bp inserts, five clones from the 12-mer library had inserts of 175 bp (one of these also had a second apparently unrelated *Eco*RI fragment of 185 bp), one clone from the 10-mer library had a 125-bp insert, and one clone from the 10-mer library had a 160-bp insert. Upon antibody screening each of the phage reacted with immune serum but did not react with preimmune serum. ³²P-labeled probes were prepared by PCR amplification (3) of one clone each from the 115-, 175-, and 125-bp size classes. The 115-bp probe hybridized with all the candidates containing 115-bp inserts and no others. Similarly, the 175-bp probe hybridized with candidates containing 175-bp inserts, and the 125-bp probe hybridized

Taq DNA polymerase (*Taq* Pol I)¹ isolated from *Thermus aquaticus* has been shown to be highly useful in the polymerase chain reaction (PCR) method (1, 2) of amplifying DNA fragments (3). The high temperature optimum activity, 75 °C, affords unique advantages when comparing *Taq* Pol I to *Escherichia coli* DNA polymerase I. High specificity of primer binding at the elevated temperature gives a higher yield of the desired product with less nonspecific amplification product. Also, *E. coli* DNA polymerase I is inactivated at 93–95 °C, the temperature range required to denature the duplex DNA product. Since *Taq* Pol I is stable at 93–95 °C, one can add

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J01639.

† To whom correspondence and reprint requests should be addressed: 1400 53rd St., Emeryville, CA 94608.

‡ The abbreviations used are: *Taq* Pol I, DNA polymerase isolated from *T. aquaticus*; kh, kilohorse(s); bp, base pairs; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; dNTP, deoxyribonucleotide triphosphate; kDa, kilodalton; X-Gal, 5-bromo-4-chloro-3-indolyl- β -D-galactoside; IPTG, isopropyl-1-thio- β -D-galactopyranoside; PBS, phosphate-buffered saline; TMB, 3,3',5,5'-tetramethyl benzidine; PCR, polymerase chain reaction; Pol I, DNA polymerase I.

§ D. Gelfand and S. Stoffel, manuscript in preparation.

¶ Portions of this paper (including "Materials and Methods," Table V, and Fig. 8) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

with only the candidate containing a 115-bp insert. Subsequent DNA sequencing of two 115-bp *Eco*RI inserts, one each from the 12-mer and 10-mer libraries, confirmed that they were identical sequences. DNA sequence analysis of *Taq* and flanking *lacZ* DNA for the candidate from the 12-mer library indicated the presence of one *Eco*RI linker at its 5' *lacZ* junction. DNA sequence analysis of the *Taq* and flanking *lacZ* DNA for the 115-bp candidate from the 10-mer library indicated the presence of three *Eco*RI linkers at the 5' *lacZ* junction, which resulted in the same frame with respect to β -galactosidase as that of the 12-mer linker candidate. Thus, we picked DNA fragments encoding the same epitope from two libraries.

Lysogens were made of all the candidates in strain Y1089 and were induced with isopropyl-1-thio- β -D-galactopyranoside (IPTG). Total proteins from crude lysates of induced cultures were run on SDS-PAGE gels, and Western blots were prepared by using the anti-*Taq* Pol I antibody for detection. All of the clones made IPTG-inducible, *lacZ*-fusion proteins which reacted with the anti-*Taq* Pol I antibody (data not shown).

One clone each from the 115-, 125-, 160-, and 175-bp insert size classes was chosen for epitope selection. This method uses crude extracts of candidate clones to select antibodies from a polyclonal antiserum. These affinity-selected antibodies were used to probe Western blots of *Taq* Pol I. The results are shown in Fig. 1. In two experiments candidate λ gt11 1, the 115-bp insert candidate, was the only one of the four tested which successfully bound antibody that reacted with purified *Taq* Pol I and reacted uniquely with *Taq* Pol I in crude extracts. The other three candidates, which had been identified and purified with the anti-*Taq* Pol I antibody, failed to "fish" from that same polyclonal antibody an antibody that would react with *Taq* Pol I on a Western blot. A close inspection of the Western blot indicates a faint cross-reaction with 28-30-kDa proteins in total soluble *Thermus* crude extracts. The DNA sequences of these three candidates do not correspond to any part of the *Taq* Pol I DNA sequence (Fig. 2).

λ Ch35 Libraries—The 115-bp *Eco*RI fragment from clone λ gt11 1 was subcloned into Genescribe Z vector pTZ19R to use as a probe in screening the λ Ch35:*Taq* library. Construction of the partial *Sau*3A digest library of *Taq* DNA in λ Ch35 and screening of the library are detailed under "Materials and Methods," in the Miniprint. The *in vitro* packaged library was plated initially on *E. coli* strain K802. That strain was chosen to avoid the possibility of degradation of *Taq* insert DNA by

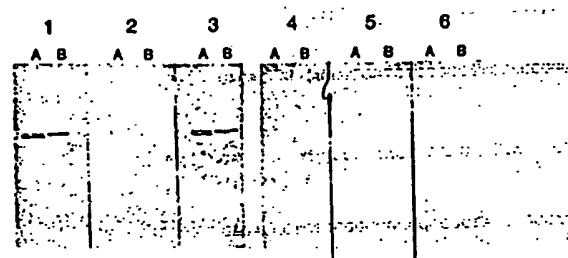


FIG. 1. Immunoblots with affinity-purified antibodies prepared by epitope selection. Epitope selection is described under "Materials and Methods." For each immunoblot, 3 units of purified *Taq* Pol I (partially proteolyzed) plus 10 μ g of gelatin were loaded on Lane A, and 10 μ g of *Taq* crude extract was loaded on Lane B. Antibodies used to probe immunoblots were: 1, 1:10,000 dilution of the anti-*Taq* Pol I polyclonal antiserum; 2, anti-*Taq* Pol I antibody affinity purified with purified β -galactosidase (negative control); 3-6, anti-*Taq* Pol I antibodies affinity purified with extracts of induced λ gt11 clones 1, 3, 9, and 2-11, respectively.

the *mcrA* or *mcrB* restriction systems (6). The amplified library was subsequently plated on *E. coli* strain MC1000.

Nine candidates were isolated and purified from the λ Ch35:*Taq* library. From restriction analysis of mini DNA preparations, none of the candidates proved to be identical, though they all shared some common restriction fragments. Upon Southern blotting, the pTZ19R probe hybridized to a common 4.2-kb *Bam*HI fragment and a common 6.5-kb *Pst*I fragment in all the candidates, consistent with the hybridization seen in Southern blots of *Taq* genomic DNA (Fig. 3). For *Hind*III, the probe hybridized to fragments of different sizes, ranging in size from 5.6 to 10 kb. In addition, all nine candidates shared a common 4.5-kb *Hind*III fragment.

One candidate, designated ϕ 4-2, had a probe-hybridizing *Hind*III fragment of approximately 8 kb which corresponded to the *Hind*III fragment that hybridized with probe I in the *Taq* genomic Southern (Fig. 3). We chose this candidate for further study and subcloned each of its four detectable *Hind*III fragments (A = 8 kb, B = 4.5 kb, C = 0.8 kb, and D = 0.5 kb) into vector BSM13* in both orientations, transforming into host DG98. The two subclones of fragment A in both orientations, pFC82.35 and pFC82.2, were IPTG-induced and extracts were assayed for *Taq* Pol I activity (Table I). Subclone pFC82.35 had IPTG-inducible thermostable activity at a very low level, which was detectable because of the high sensitivity of the assay (<1 molecule/10 cell equivalents). In contrast, pFC82.2 had a significantly lower basal level of *Taq* Pol I activity which was attenuated in extracts of IPTG-grown cultures.

A restriction map of the A fragment was generated and is shown in Fig. 4. Southern analysis showed that the λ gt11 1 probe hybridized at one end of the A fragment. Indeed, the DNA sequence of the *Alu*I genomic fragment isolated in λ gt11 1 corresponds to nucleotides 619-720 in the *Taq* Pol I gene (Fig. 2). Further, the *Eco*RI-adapted *Alu*I site at the junction between *E. coli* *lacZ* and *Taq* in λ gt11 1 corresponds to the *lac* promoter-proximal *Taq* *Hind*III site in pFC82.35.

*Deletions in the A Fragment to Localize the *Taq* Pol Gene*—Two different deletions were made in the A fragment in pFC82.35 to aid in localizing the gene. In pFC84, approximately 2.4 kb of the right end of the A fragment was deleted from the *Sph*I site (Fig. 4) rightward to the *Sph*I site in the vector polylinker. In pFC85, approximately 5.2 kb of the right end of the A fragment was deleted from the *Asp*718 site rightward (Fig. 4) to the *Asp*718 site in the vector polylinker, leaving 2.8 kb of *Taq* insert sequence. The activity of *Taq* Pol I was assayed in extracts of uninduced and IPTG-induced pFC84 and pFC85 in DG101. As can be seen in Table I, deleting 3' sequences in the A fragment had a dramatic effect on the IPTG-inducible expression of *Taq* Pol I. In addition, while we were unable to detect *Taq* Pol I in Western blots of IPTG-induced pFC82.35/DG98, induced immunoreactive bands were clearly seen upon Western blotting of IPTG-induced pFC84/DG101 and pFC85/DG101 (Fig. 5). In the Western blots, induced pFC84/DG101 and pFC85/DG101 lanes revealed doublet immunoreactive bands that were approximately 65- and 63-kDa. These immunoreactive species were considerably smaller than full-length 94-kDa *Taq* Pol I. We determined that the doublet bands were not artifacts of the gel analysis because they were seen repeatedly in several experiments.

LacZ α Fusions—To define further the locus of the *Taq* Pol I gene and to confirm the reading frame at different sites for use as guideposts during DNA sequence analysis, we constructed several fusions of the left end of the *Taq* *Hind*III A fragment to *lacZ α* in the BSM13* vector. These fusions are

FIG. 2. DNA sequence and deduced amino acid sequence of the *Taq* *PoI* I gene. Nucleotides were numbered consecutively from the start of the gene. Nucleotide numbers are shown on the left. Amino acid numbers are shown on the right.

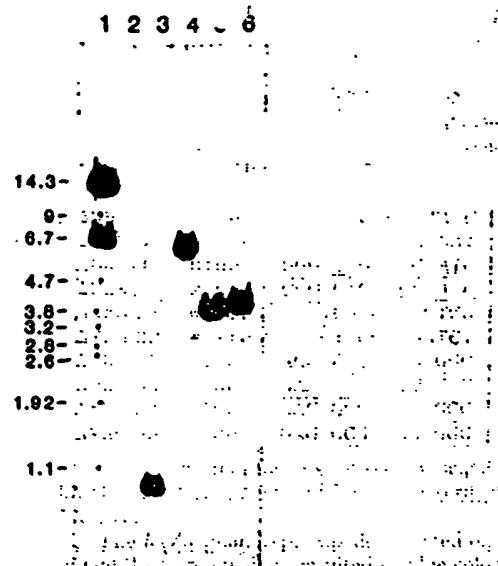


FIG. 3. Southern blot analysis of *Taq* genomic DNA probed with $\alpha^{32}\text{P}$ -labeled PCR-amplified probe. Lane 1 is a size standard *Eco*RI- and *Bam*HI-digested λ plac and *Msp*I-digested plasmid Lac5. DNA fragment sizes (in kilobases) are listed at left. The PCR-amplified probe contains the *λgt11* primer sequences on either end (flanking the *Eco*RI site in *lacZ*) which are homologous to sequences in the 14,300 and 6,700 marker bands. Lanes 2-6 are *Taq* genomic DNA digested with *Hind*III, *Hind*III and *Pst*I, *Pst*I, *Pst*I and *Bam*HI, and *Bam*HI, respectively.

TABLE I
Taq DNA polymerase activity in *E. coli* extracts

Experiment	Extract	IPTG	Specific activity ^a
I	BSM13*	±	<0.001 ^b
	BSM13* <i>w/Taq</i>	+	0.142
	BSM13* <i>w/Taq</i>	+	0.136
	pFC82.35	-	0.248
	pFC82.2	+	0.310
II	BSM13*	+	0.003 ^c
	pFC84	-	1.24
	pFC85	-	29.7
	pLSG1	+	0.87
	pLSG1	-	29.6

^a Specific activity in units/mg total crude extract protein when assayed, as described under "Materials and Methods," on clarified, heat-treated extracts.

^b A background of 0.004% input counts has been subtracted. Extract protein corresponding to 3×10^7 cells was assayed.

^c Purified *Taq* DNA polymerase was added to a replicate cell pellet at time of lysis. The assay contained 4×10^7 molecules of *Taq* Pol I.

^d Purified *Taq* Pol I, corresponding to 4×10^7 molecules, was admixed with the BSM13* extract at time of assay.

^e A background of 0.002% input counts has been subtracted. BSM13* specific activity represents two times background.

described under "Materials and Methods" and are summarized in Table II. Using these fusions we determined the reading frame of *Taq* Pol I at the *Nhe*I site at nucleotide 2043, the *Bam*HI site at nucleotide 1780, and at four locations at or leftward of the *Xba*I site at nucleotide 1408.

Assembling the Full-length *Taq* Pol I Gene—As described above, the *Sph*I and *Asp*718 deletants, pFC84 and pFC85, produced thermostable polymerase activity upon induction. However, the size of the induced bands detected by anti-*Taq* Pol I antibody in Western blots was smaller than full-length

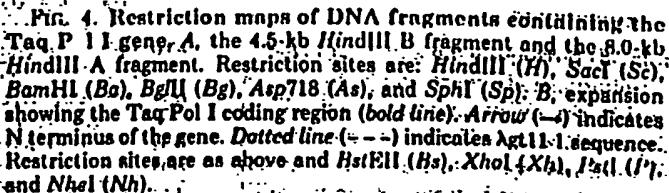


FIG. 4. Restriction maps of DNA fragments containing the *Taq* Pol I gene. *A*, the 4.5-kb *Hind*III B fragment and the 8.0-kb *Hind*III A fragment. Restriction sites are: *Hind*III (H), *Sac*I (Sc), *Bam*HI (Ba), *Bgl*II (Bg), *Asp*718 (As), and *Sph*I (Sp). *B*, expansion showing the *Taq* Pol I coding region (bold line). Arrow (→) indicates N-terminus of the gene. Dotted line (---) indicates *λgt11-1* sequence. Restriction sites are as above and *Hst*III (Hs), *Xba*I (Xb), *Pst*I (Ps), and *Nhe*I (Nh).

polypeptides. The size of the *Taq* gene product was determined by SDS-PAGE analysis of the *Taq* Pol I clones. The *Taq* Pol I clones were induced with IPTG as described under "Materials and Methods." Uninduced and induced samples were analyzed on SDS-PAGE gels and subjected to Western blot analysis, also as described under "Materials and Methods." Lane 1, *λCh35* low molecular weight marker. Molecular weights (in thousands) are shown at left. Lane 2, induced BSM13* (33 μ g) negative control. Lanes 3 and 4, uninduced (0.04 unit, 33 μ g) and induced (1.0 unit, 33 μ g) pFC85. Lanes 5 and 6, uninduced (0.03 unit, 33 μ g) and induced (1.0 unit, 33 μ g) pFC84. Lanes 7 and 8, uninduced (0.05 unit, 11 μ g) and induced (0.4 unit, 11 μ g) pLSG1. Lane 9, BSM13* (33 μ g) plus 0.4 unit of purified *Taq* Pol I.

FIG. 5. Western blot analysis of *Taq* Pol I clones. Cultures of *Taq* Pol I clones were induced with IPTG as described under "Materials and Methods." Uninduced and induced samples were analyzed on SDS-PAGE gels and subjected to Western blot analysis, also as described under "Materials and Methods." Lane 1, *λCh35* low molecular weight marker. Molecular weights (in thousands) are shown at left. Lane 2, induced BSM13* (33 μ g) negative control. Lanes 3 and 4, uninduced (0.04 unit, 33 μ g) and induced (1.0 unit, 33 μ g) pFC85. Lanes 5 and 6, uninduced (0.03 unit, 33 μ g) and induced (1.0 unit, 33 μ g) pFC84. Lanes 7 and 8, uninduced (0.05 unit, 11 μ g) and induced (0.4 unit, 11 μ g) pLSG1. Lane 9, BSM13* (33 μ g) plus 0.4 unit of purified *Taq* Pol I.

Taq Pol I, i.e. approximately 65-kDa as opposed to full-length 94-kDa. Thus, we felt that the A fragment lacked the 5' portion of the gene which would encode the N-terminus.

Also mentioned earlier, all candidates from the *λCh35* library which had been identified with the pTZ19R:1 probe shared a common, approximately 4.5-kb *Hind*III fragment which did not hybridize to the probe. This fragment, the B fragment, was subcloned into BSM13*, yielding plasmid pFC83. The restriction map of the B fragment was determined (Fig. 4). By comparing those mapping results and the A fragment map with the results of *Taq* genomic Southern blots probed with probe 1 (Fig. 3) we deduced that *Hind*III fragment B was likely to contain the 5' portion of the *Taq* Pol I gene.

The 724-bp *Bgl*II-*Hind*III segment of the B fragment was subcloned into *Bam*HI- and *Hind*III-digested BSM13*. Upon sequencing, an ATG and subsequent open reading frame was found 109 bp downstream of the *Bgl*II site. The open reading frame continued to the *Hind*III site. In addition, the phase of the open reading frame at the "right" end of the B fragment was identical to the phase of the open reading frame at the "left" end of the A fragment.

PCR amplification confirmed that the B and A fragments in pFC83 and pFC82.35 are contiguous in the *Taq* genome. Primers were chosen which flanked the presumed internal *Hind*III site: MK138 (Table V, in the Miniprint) on the left side of *Hind*III and FL25, a 20-mer complementary to nucleotides 622-641 of the *Taq* Pol I sequence, on the right side of *Hind*III. Upon amplification (3) of the *λCh35* genomic phage

TABLE II
LacZα fusions

Fusion*	LacZα phenotype ^b	Fusion DNA sequence ^c		Polylinker A
		Taq	BSM13	
ΔNhe 1	Blue	<u>GAG</u> CTA' G	<u>CG</u> AGC TCG	lacZα start
ΔBa 15	White	<u>CAGAGGAT</u>	<u>CCC</u> CGG <u>GTA</u>	
ΔBa 33	Blue	<u>GGG</u> CAG ARG ATC	<u>GAT</u> CCC <u>CGG</u> GTA	
ΔBa 35	Blue	<u>GGG</u> CAG AGG ATC	<u>CCC</u> CGG <u>GTA</u>	
ΔXh 28	White	TCGCCCGCCTCG	GTA CGG AGC TCG	
ΔXho 30	White	GTGGCCGGATCT	TA CGG AGC TCG	
ΔXho 32	Blue	AGG CTT GAG GGG	GTA CGG AGC TCG	
ΔXho 53	Blue	GAA GGC CTT GGC	GTA CGG AGC TCG	
ΔXho 54	Blue	GAG GGG ATG GCC	CCG AGC TCG AAT	
ΔXh 59	Blue	GAG GCG CGG GGG	GTA CGG AGC TCG	

* Construction of fusions between 5' sequences of the *Taq Pol I* 8.0-kb *Hind*III A fragment and *lacZα* is described under "Materials and Methods."

^b The *lacZα* phenotype was determined on agar plates containing X-Gal. In-frame fusions resulted in blue colonies on X-Gal and out-of-frame fusions yielded white colonies.

^c The DNA sequence was determined at the site of each fusion. BSM13⁺ polylinker sequence is shown to the right of the bold line. Groupings of three nucleotides indicate the reading frame of *lacZα*. *Taq* DNA sequence is shown to the left of the bold line. For in-frame (blue) fusions, the deduced frame of the *Taq Pol I* gene is indicated. Restriction sites regenerated (*Nhe*1, *Bam*H1) or generated (*Clal*) are indicated by italics. The *Taq Pol I* nucleotide coordinates (Fig. 2) at the fusion sites of the *Xba*I *lacZα* fusions are: ΔXho 28, 1411; ΔXho 30, 962; ΔXho 32, 1266; ΔXho 53, 1098; ΔXho 54, 1173; ΔXho 59, 1050.

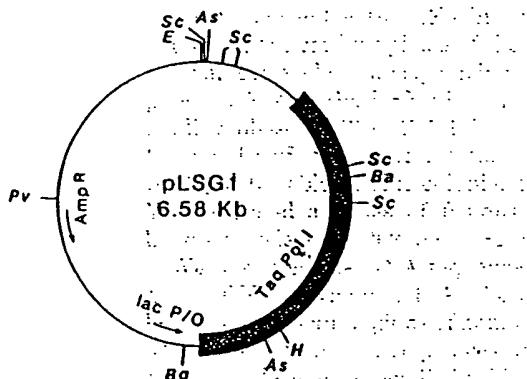


FIG. 6. Plasmid pLSG1. The 6.58-kb plasmid contains a 3.41-kb segment of *T. aquaticus* DNA in a derivative of the plasmid vector BSM13⁺. The bold line indicates the 2.5-kb *Taq Pol I* coding sequence. Expression of *Taq Pol I* is controlled by the *lac* promoter/operator. Construction of the plasmid is described under "Materials and Methods." Restriction sites are as in legend to Fig. 4 and *Eco*RI (*E*) and *Pvu* (*Pv*).

φ4-2, we observed the predicted PCR product of 86 bp (data not shown), indicating that the B and A *Hind*III fragments are contiguous. A larger PCR product would have indicated that there was another *Hind*III fragment in the gene.

The assembly of the full-length *Taq Pol I* gene in plasmid pLSG1 (Fig. 6) is described under "Materials and Methods." Cultures of pLSG1 in DG101 produced IPTG-inducible thermostable polymerase activity at 37.5 units/mg protein (Table I). Western blots of IPTG-induced pLSG1/DG101 cultures revealed an immunoreactive band of approximately the same size as full-length *Taq Pol I*, 94-kDa (Fig. 5). Coomassie staining of IPTG-induced pLSG1/DG101 cultures failed to indicate the presence of a detectable induced band. The complete nucleotide sequence of the *Taq Pol I* gene and the deduced amino acid sequence are presented in Fig. 2. The DNA sequence of the *Taq Pol I* gene predicts an open reading

frame of 2496 bp with a G + C content of 67.9%. The DNA sequence AAGG (-9 through -6, Fig. 2) is complementary to the 3' end of *E. coli* and *Thermus thermophilus* 16 S rRNA (7) and may comprise a portion of the ribosome binding site for initiation of translation at the first ATG.

DISCUSSION

Several groups have reported the cloning and expression in *E. coli* of genes from thermophiles: malate dehydrogenase (*mdh*) from *Thermus flavus* (8), β -isopropylmalate dehydrogenase (*leuB*) from *Thermus thermophilus* (9), and the *TaqI* restriction-modification system from *Thermus aquaticus* (10). Iijima et al. (8) selected the *mdh* gene from a *T. flavus* partial *Hind*III library in pBR322 by screening crude extracts of pools of independent library transformants at 60 °C for malate dehydrogenase activity. Nagahashi et al. (9) selected directly for expression of the *leuB* gene in *E. coli*. Although the activity of the enzyme at 37 °C was quite low compared to its activity at 75–80 °C, they were able to recover clones which complemented a *leuB* mutation in the *E. coli* host. Slatko et al. (10) also selected directly for expression of *TaqI* methylase in *Taq*:pBR322 libraries. However, *TaqI* endonuclease appeared not to be active at 37 °C in *E. coli*, since clones with only the restriction gene were viable in the absence of modification.

Several groups have also reported cloning and expression of DNA polymerases in *E. coli*. Kelley et al. (11) cloned the structural gene for DNA polymerase I (Pol I) from *E. coli* in λ bacteriophage. They observed polymerase activity in the transducing phage at a level of approximately 4% of total cell protein. However, they were unable to maintain a plasmid harboring the *Pol I* gene, probably because overproduction of *Pol I* in *E. coli* is lethal to the cell. More recently, T4 DNA polymerase has been cloned and expressed in *E. coli* (12). In this case, it was necessary to clone the gene under control of inducible promoters such that constitutive expression of the gene would be minimal. Attempts to clone the gene under control of its own promoter in *E. coli* were unsuccessful, probably because of the detrimental effect the polymerase had on the cell. We did not know if *Taq Pol I* would be toxic to *E. coli* cells at 37 °C. While the *in vitro* specific activity of *Taq Pol I* at 37 °C is only a few percent of the specific activity at 75 °C, we could not predict if the DNA binding activity of the enzyme might interfere with normal cell function. To avoid potential problems related to direct expression of the gene in *E. coli* we chose to clone an epitope of the *Taq Pol I* gene by using λ gt11 libraries and antibody selection. The epitope-expressing clone was subsequently used to select the entire *Taq Pol I* gene from a library in λ Ch35.

We were unable to detect a thermostable polymerase activity in cells infected (11) with any of the λ Ch35 clones, including φ4-2. The polymerase assay is extremely sensitive and can detect 1 molecule of polymerase per 10 cell equivalents. Upon subcloning of the 8-kb probe-hybridizing *Hind*III A fragment from φ4-2 into BSM13⁺ and IPTG induction of the subclone pFC82.35, a low level of thermostable polymerase activity was detected (Table I). Based on the activity of purified *Taq Pol I* when admixed with *E. coli* cells, this activity represents two to three molecules of *Taq Pol I* per cell equivalent. The gene was localized to one end of the 8.0-kb *Hind*III A fragment by using deletion analysis. Upon IPTG induction, pFC84, the *Sph*I deletion, and pFC85, the *Asp*718 deletion, yielded a 100-fold increase in *Taq Pol I* activity (Table I) compared to that of the full-length A fragment subclone, pFC82.35. This increase in activity allowed for ready detection of the induced protein(s) on Western blot (Fig. 5). The A fragment induced proteins were truncated with an

apparent molecular mass of 63-65 kDa.

Fusing the 5' *Hind*III site in the A fragment with the *Hind*III site in BSM13* causes the *Taq Pol I* gene to be out of frame with respect to β -galactosidase. The reading phase at the *Hind*III site in BSM13* with respect to β -galactosidase is A AGC TT, a frame of "0" (13). The reading frame of *Taq Pol I* at the *Hind*III site is AAG CTT ("plus 1"). The fusion gives rise to a minus 1 frame shift. In the β -galactosidase reading frame, there is a TGA stop codon at nucleotide 1478 of *Taq Pol I*. Downstream of this TGA there are several possibilities for restarts which could result in truncated forms of *Taq Pol I*: ATGs at nucleotides 1509 and 1752 and GTGs at nucleotides 1547, 1569, 1722, and 1731. In fact, we see a doublet in induced lanes of both pFC84 and pFC85 on Western blots (Fig. 5) indicating at least two reinitiation sites. All but one of the likely sites, the ATG at nucleotide 1509, would probably require a ribosome binding site for reinitiation. There are reasonable ribosome binding sites for the GTG at nucleotide 1722 and for the ATG at nucleotide 1752. Translation initiating at these sites would yield proteins of 59 and 58 kDa, respectively. However, the apparent molecular masses of the doublet bands seen on Western blots of pFC84 and pFC85 are approximately 65 and 63 kDa, based on comparison of the mobilities of the doublet bands with the molecular weight size standards. Whether the result of reinitiation or proteolytic processing, the thermostable, enzymatically active, truncated forms of *Taq Pol I* directed by plasmids pFC84 and pFC85 (Table I) suggest that significant portions of the *Taq Pol I* sequence are not essential for DNA polymerase activity.

The purpose of the set of fusions of 5' portions of the *Taq Pol I* A fragment with *lacZ* in BSM13* was to confirm or determine the reading phase of the *Taq Pol I* gene internally as an aide to nucleotide sequencing. Since we knew the reading phase of *lacZ* in the BSM13* polylinker, we could infer the reading phase of *Taq Pol I* in α -complementing in-frame fusions. DG98 harboring fusions which were in-frame were readily detectable as blue colonies on *X*-Gal indicator plates. We generated a series of fusions (Table II) at nine sites between nucleotides 962 and 1782 of the *Taq Pol I* gene.

We compared the DNA sequence of *Taq Pol I* with that of *E. coli* DNA polymerase I. At the DNA level, the two genes lack any significant regions of homology (Table III). In regions where the amino acid sequences are homologous, the DNA sequences diverge, especially in third positions of codons. The longest stretch of DNA sequence identity is 19 bases (Table III).

The predicted amino acid sequence of *Taq Pol I* is shown in Fig. 2. From this a codon bias table was generated (Table IV). There is a heavy bias toward G and C in the third position (91.8% C and G) as would be expected for GC-rich organisms.

TABLE III
DNA sequence identity of *Taq Pol I* and *E. coli* *Pol I*

	Sequence location*	Nucleotide identity	Amino acid identity
<i>Taq Pol I</i>	190-208		
<i>Pol I</i>	178-196	19/19	6/6
<i>Taq Pol I</i>	1730-1757		
<i>Pol I</i>	2015-2042	23/28	9/9
<i>Taq Pol I</i>	2260-2277		
<i>Pol I</i>	2545-2562	17/18	6/6
<i>Taq Pol I</i>	2344-2363		
<i>Pol I</i>	2635-2654	17/20	7/7

* Nucleotide sequence coordinates for *Taq Pol I* from Fig. 2. Nucleotide sequence coordinates for *E. coli* *Pol I* adapted from GenBank.

TABLE IV

Codon usage in the *T. aquaticus* DNA polymerase I gene

Arg	CGT	0	Leu	TTG	3	Ser	TCT	0
(76)	CGC	24	(124)	TTA	0	(31)	TCC	16
	CGG	27		CTT	20		TCG	1
	CGA	0		CTC	46		TCA	0
	AGG	25		CTG	60		AGT	1
	AGA	0		CTA	5		AGC	14
Thr	ACT	0	Tyr	CCT	3	Val	GTC	1
(30)	ACC	20	(48)	CCC	34	(51)	CTC	21
	ACG	10		CCG	9		GTG	29
	ACA	0		CCA	2		GTA	0
Ala	GCT	2	Gly	GTC	0	Ile	ATT	3
(91)	GCC	77	(68)	GGC	28	(25)	ATC	20
	GCG	12		GGG	30		ATA	2
	GCA	0		GGA	0			
Asn	AAT	0	Gln	CAG	15	Tyr	TAT	4
(12)	AAC	12	(16)	CAA	4	(24)	TAC	20
His	CAT	0	Gl	GAG	79	Cys	TCT	0
(18)	CAC	18	(87)	GAA	8	(0)	TGC	0
Asp	GAT	3	Phe	TTT	8	Lys	AAG	37
(42)	GAC	39	(27)	TTC	19	(42)	AAA	5
Met	ATG	16	Trp	TGG	14			
(16)					(14)			

and as others have observed for other *Thermus* genes: 95% C and G for the *gk24* gene encoding L-lactate dehydrogenase of *Thermus caldophilus* (15), 94.8% for *mdh* from *T. flavus* (14), and 89% for *leuB* from *T. thermophilus* (16).

Significant amino acid sequence similarity exists between *Taq Pol I*, *E. coli* *Pol I*, and bacteriophage T7 DNA polymerase. One possible sequence alignment yields 38% identity between the *Taq Pol I* and *E. coli* *Pol I* amino acid sequences (Fig. 7). There are two major regions of *Taq Pol I* and one region of T7 DNA polymerase that show extensive sequence similarity compared to *E. coli* *Pol I*. The first region of *Taq Pol I* extends from the N terminus to approximately residue 300. The second region extends from approximately residue 410 to the C terminus of *Taq Pol I*. The N-terminal region of *Taq Pol I* corresponds to the N-terminal domain of *E. coli* *Pol I* shown to contain the 5'-3' exonuclease activity (17). The C-terminal regions of *Taq Pol I* and T7 DNA polymerase correspond to the *E. coli* *Pol I* domain shown to contain DNA polymerase activity (18). The x-ray structure of the Klenow fragment (19) shows that this domain contains a deep cleft believed to be responsible for DNA binding.

Apparently as a result of many mutations, deletions, insertions, etc. during evolution, *Taq Pol I* residues at positions 300-410 show little sequence similarity compared to *E. coli* *Pol I*. *Taq Pol I* is 96 residues shorter than *E. coli* *Pol I*; most of the deleted residues occur in the region encompassing residues 300-410. Ollis et al. (19) and Derbyshire et al. (20) have shown that *E. coli* *Pol I* residues Asp-355, Glu-357, Leu-361, Asp-424, Phe-479, and Asp-501 are involved in binding of divalent cation and deoxyribonucleoside monophosphate. A fragment of *E. coli* *Pol I* that contains only residues 515-928 is devoid of 3'-5' exonuclease activity, but still retains polymerase activity (18). Presumably, the *E. coli* *Pol I* region comprised of residues 324-515 forms at least part, if not all, of the 3'-5' exonuclease activity. *Taq Pol I* and *E. coli* *Pol I* display little sequence similarity in the presumptive 3'-5' exonuclease region. Of the *E. coli* *Pol I* residues shown to be involved in cation and deoxyribonucleoside monophosphate binding, the sequence alignment of Fig. 7 shows only Asp-424 as having an exact homolog in the *Taq Pol I* sequence. Although other high scoring sequence alignments are possible in the *Taq Pol I* 300-410 region, it is possible that the *Taq Pol I* gene has undergone key mutations, deletions, or insertions

T.aq.	1	MRCGLPLFEPKGAVVLLVGDIIIAYRTFHAKGLTTSRGEVQ...AKSLKAKLZ...DCDAVIVVFDAKSPRHEATCGGYKAGRPTPQDFPRLALI...	99
E.c.	1MVQTPQNPFLILVQDGESELYRAYHAPPNTNSAGEPTG...PLAHLRSLSINGYKPTHAAYVFDACKTFRQELFEHYKSHAPPMDDLRQYQEL...	95
T.aq.	100	KELVDLILGLARLLEVPGYEADVLASLAKRAKEKEGTEVRLTADKDLQYOLLSDRIVHLHP...EGYLITPAULWEEKYGLRPDQHADYRALTGDESDNLPGVKG...	198
E.c.	96	HAMVKAHGLPLLAWSGVDEADVIGTILAREAEATRACRPPVLESTGDKDQHAEQVTPHITLINTHTLGPVEVNNKYGVPPELITDPLALMGDSSDNIPGVKG...	195
T.aq.	199	IGEKTARKLLEEMGSLEALLKLNDL...KPAIREKELAKHODDLKLSHDILAKVTDPLPEVDFAK...RPREPREKRAFLEP...	280
E.c.	196	VGEKTAQALQLOGLGQDLYAEPKIAQESFRGAXTHAAKLEQNKEVATLSTOLATXKTOVELETCTCQDQLEVOQPALEBLLGLKRTKEKQKNTADVZAGK...	295
T.aq.	281	LLHEFGLLESPKALEEA...PWPPIPGEVLSRKEPHWADILALABARGGRVRAPEPYKAL...RDLKEARGL...	351
E.c.	296	WLQAKGAKPAAKPQETSV...DEAPEVTATVISXONYVTILDEETLKAWIAKLEKAFVFAFDTETOSLDNISANLVGLSFAIEGVAAYIPVARDYLDAPDQ...	395
T.aq.	352	LAQDLSQLALREGGLPPI...DDPMILAYLLOPSNTP...EGVARRYGGENTEEAGERAALSER...	411
E.c.	396	ISRERALELLKPLLEDERALKVQONLKYORGILANYGIELRGIAF...DTMLESYILNSVAGRHDMSLAEPHMKHKTTFEEIACKGKQNLTFN... The above sequence is similar to the sequence in the following row.	487
T.aq.	412	...LPAHNGRLEGEEBLILYREVERPLSAVLAHMEATGVRLDVAYLRAISLEVAEEETRLEAEVFRLAGHOFPNLNSRDQLE...	491
E.c.	488	QIALEEEAGRYAREADAVTLQLHLKQWPDQKHKGPILNFEENIEMPLVPLSRIERNVKIDPKVLUHNSEELTLRLELEKKAHBIAGEEFNLSSTKQQL...	587
T7	334	...FNPSSSRDHIQ...	343
		570	
T.aq.	492	RVLFDELGLPAIGKTEKTGKRSTSAAVLEALR...EAPHPIVEKILQYRELTKLKSTYDPLPDLH...PRTGRHLTRENTATATGRLSSSDPNLQNIF...VR...	587
E.c.	588	TILFEKGQIKPLKKT...PGGPSTSEEVLEELA...LDYPLPKVILEYRGGLAKLKSTYDPLPLH...PKTGBVHTSXHQAATPTGRILSSDTPNLQNIF...VR...	682
T7	343	KKLQE...AGWVPTKYTDK...GAPVVDDEVLEGVRVDDPEKQAAIDQIKEYLH1QKRIGQSAEGDKAHLYVAEDGKIHGSVPNGAVTGRATHAEPNLAQIPGVR...	444
T.aq.	588	TPLGORIRRAFIAE...EGHLVALDYSQIELRVLALHSGDENLIRVFOEGRDINTETASW-MFGVPREAVDPLNRRAAATINFGL...GMSAHLRQLSQEL...	682
E.c.	683	NEEGRRIQAFIAP...EDYVYVSADYSQIELRIMAHLSRDKGLLTAAPEKDINRATALEVFGLPLETVSEQRSLAATINFGL...GMSAFGLARQL...	777
T7	445	SPYGEQCRAFGAEHHLDGITGKPHWQAGIDASGLELRLCAH...FMAFEDNGEYAHETLNGDIHTKHNQIAEELPTRDNAATFIYGFELIGAGDEKIGQIV...	541
T.aq.	683	AIPYEEAQAFIERYFQSFPKV...RAHIEKTLEEGRRFQYETLFG...RRRYVPDLEARVKSvreAAERHAFNHPVQGTAADLMKLLAMVKIFPRL...EEMG...	776
E.c.	778	NIPRKEAQKYMOLYFERYPGV...LEYHMRTRAQAKEQYETLDG...RRLYLPDKISSNGARRAMAERAIAHAPQGTAADITKRAHIAVDAAHLQAEQPR...	873
T7	542	GAGKERGKELKKFLENTPAIAALRESTIQQTLVESSQWAGEQQVKMKRHNKGLDGRKVHVRSP...HAALINTLQSGALICKLWIKTEEMLVEGLK...	639
T.aq.	777	A.....RMLLQV...DELVLEAPKER...AEAVARIAKEVM...EGVYPLAVPLVEVGVIGEDHLSAKE...	832
E.c.	874	V.....RHIMQV...DELVVEEVHKDD...VDAVAKQIHQML...ENCTRLDVPPLLVEVGSGENNDQAH...	928
T7	640	HGHGDDFAYHAWBDEIOVGCRTEETAGVVIETAOEAMHRHVGDHWNFRCLLDTEGKQGPNNHAICH...	704

FIG. 7. Amino acid sequence comparison of the DNA polymerases from *T. aquaticus*, *E. coli*, and bacteriophage T7. Deduced amino acid sequences for Taq Pol I (*T.aq.*), *E. coli* Pol I (*E.c.*), and bacteriophage T7 DNA polymerase (*T7*) were analyzed for amino acid sequence homology by using the computer program GAP from the University of Wisconsin Genetics Computer Group. The alignment was obtained by using the mutational data scoring matrix of Staden (70). Vertical marks denote amino acid identities or functional relatedness between pairs of residues in the three sequences. Half-vertical marks denote amino acid identities or functional relatedness between residues in Taq Pol I and T7 DNA polymerase.

that have destroyed its 3'-5' exonuclease activity. Preliminary results indicate that Taq Pol I displays little if any 3'-5' exonuclease activity.

Sequence homology between *E. coli* Pol I and T7 DNA polymerase has been previously noted. Those T7 DNA polymerase sequences shown by Ollis *et al.* (21) to be conserved between that enzyme and *E. coli* Pol I are also present in the Taq Pol I amino acid sequence (Fig. 7). Most of the conserved residues are found in structural features that form the DNA-binding cleft of the enzyme. Although short segments of T7 DNA polymerase sequence in the 1-334 region are similar to

domain, is poor. A complete and unambiguous sequence alignment for this region cannot be assigned. It should be noted that although T7 DNA polymerase also shows little similarity to *E. coli* Pol I in the region of the 3'-5' exonuclease domain, T7 DNA polymerase is reported to display significant 3'-5' exonuclease activity (22, 23).

Bernad *et al.* (24) and Pizzagalli *et al.* (25) have identified several short regions of DNA polymerase amino acid se-

quences that are highly conserved. The conserved sequences are found in polymerases from herpes simplex virus type 2, human cytomegalovirus, varicella-zoster virus, Epstein-Barr virus, vaccinia virus, adenovirus type 2, killer plasmid from *Klebsiella lactis*, maize mitochondrial particle, bacteriophage ϕ 29, bacteriophage T4, bacteriophage PRD1, and yeast plasmids. Neither *E. coli* Pol I, Taq Pol I, nor bacteriophage T7 DNA polymerase contains the conserved sequences noted in the polymerases from those sources. Aside from the homology between Taq Pol I and either *E. coli* Pol I or T7 DNA polymerase, no significant amino acid sequence similarity was found when a global homology search was made comparing Taq Pol I to the National Biomedical Research Foundation's amino acid sequence database.

Chemical modification and inactivation studies of *E. coli* Pol I have resulted in the identification of many amino acid residues believed to be important or essential for polymerase activity (26-31). Among these residues are: Met-512, Arg-682, Lys-758, Tyr-766, Arg-841, and His-881. Comparing the Taq Pol I amino acid sequence to the *E. coli* Pol I sequence, all of the above residues, except Met-512, are conserved. Taq Pol I contains a Leu residue at the analogous position. Apparently, the functionally similar Taq Pol I Leu residue at position 417 can fulfill the role ascribed to *E. coli* Pol I Met-512 in template primer binding (30).

Analyses of the effects of various mutations in the *E. coli* Pol I gene upon enzymatic activity have also been used to define amino acid residues important for polymerase activity. For example, a Gly to Arg mutation at position 850 (*polA5*) results in a polymerase that is less processive on the DNA substrate (32). An Arg to His mutation at position 690 (*polA6*) results in a polymerase that is defective in DNA binding (33). Both Gly-850 and Arg-690 are conserved residues in Taq Pol I. Joyce *et al.* (34) have characterized a number of *E. coli* Pol I mutants defective in 5'-3' exonuclease activity. Interestingly, the four mutations, Y77C (*polA107*), G103E (*polA413*), G184D (*polA480ex*), and G192D (*polA214*) all occur at amino acid residues that are conserved in Taq Pol I.

As would be expected for an enzyme from a thermophilic organism, Taq Pol I is considerably more thermostable than Pol I from *E. coli* (data to be presented in a later publication). Although a better assessment of an enzyme's thermostability would result from a complete cataloging of all stabilizing amino acid interactions, in the absence of high-resolution x-ray crystal structures, many researchers have attempted to explain enzyme thermostability by an analysis of amino acid content (35-37). Several features of thermostable enzymes have been noted in such studies. Among those features are increased ratios of Arg to Lys residues, Glu to Asp residues, Ala to Gly residues, Thr to Ser residues, and a reduced Cys content. Comparing Taq Pol I to *E. coli* Pol I, the Ala to Gly and Thr to Ser ratios are smaller for Taq Pol I than for *E. coli* Pol I. Of the thermostabilizing type amino acid alterations that hold true, it is particularly notable that the Arg-to-Lys ratio for Taq Pol I is nearly twice that for *E. coli* Pol I. It is possible that the propensity of thermophilic proteins to contain Arg rather than Lys residues is simply a reflection of the high GC content of thermophilic organisms. The structural gene for Taq Pol I contains 67.9% GC compared to a 52.0% GC content for *E. coli* Pol I. The six Arg codons are rich in G and C (13 out of 18 bases are G or C) compared to the two Lys codons (1 out of 6 bases is a G). This explanation for amino acid preferences in proteins from thermophilic organisms cannot be the basis for Glu versus Asp, Thr versus Ser, or Ala versus Gly preference, because there are equal ratios of GC versus AT in the codons for those pairs of amino acids.

A more likely explanation for the preference for Arg over Lys in thermostable proteins would seem to be based on the unique physical-chemical properties of the two amino acids (e.g. pK_a values, hydrogen bonding patterns, hydrophobicity/hydrophilicity).

The truncated and full-length Taq Pol I enzymes produced upon IPTG induction show different reactivities to the anti-Taq Pol I antibody. For Western blots (Fig. 5), the immunoreactive band in the lane of induced pLSG1 is more readily detectable than induced pFC84 or pFC85, the *Sph*I and *Asp*718 A fragment deletions. In fact, we loaded three times as much of the pFC84 and pFC85 extracts compared to pLSG1, and the resulting pLSG1 immunoreactive band is still more intense. We infer that there are more epitopes for our antibody, prepared from full-length (94-kDa) Taq Pol I SDS-PAGE gel slice, in the N-terminal end of Taq Pol I than in the C-terminal two thirds of the protein. Or, based on activity, there is at least a 3-fold difference in reactivity with the antibody of the truncated versus the full-length form of the enzyme.

The level of expression in *E. coli* of full-length Taq Pol I encoded by pLSG1 is similar to the level of expression of Taq DNA polymerase in *T. aquaticus*. In pLSG1 (Fig. 6), the beginning of the Taq Pol I open reading frame is 109 bp distal to the *Bgl*II site and 171 bp distal to the *lacZα* translation initiation site. A low level of Taq Pol I expression in cells harboring pLSG1 is consistent with an in-phase TGA codon (-111 through -109, Fig. 2) in the Taq DNA sequence causing translation termination of the *lacZα* polypeptide. Reinitiation of translation at the first ATG results in the synthesis of the 94-kDa Taq Pol I protein. Further manipulation of the Taq DNA polymerase sequence has increased the level of expression. The cloned full-length Taq Pol I gene in pLSG1 affords the advantages of expressing Taq Pol I in *E. coli* and in ease of isolating the enzyme from *E. coli* compared to *T. aquaticus*. These advantages will aid in further study of the enzyme and will provide a ready source of Taq Pol I for use in PCR and other biochemical procedures in which Taq Pol I might prove useful, such as in DNA sequencing.

Acknowledgments—We gratefully acknowledge Gail Rodgers for advice on preparation of antibody from a small amount of protein; Corey Levenson, Lauri Goda, and Dragan Spasic for preparation of oligonucleotide primers, Keith Bauer for fermentation support; Will Bloch and David Birch for advice on nonradioactive detection of proteins; Henry Erlich, Tom White, John Sninsky, and Michael Innis for advice and support; Hamilton Smith and Norman Arnheim for advice and critical review of the manuscript; Sharon Nilson and Eric Ladner for preparation of figures; and Patricia A. Robinson and Edna McCallan for preparation of the manuscript.

REFERENCES

1. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350-1354.
2. Mullis, K. B., and Faloona, F. A. (1987) *Methods Enzymol.* **155**, 335-350.
3. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487-491.
4. Chien, A., Edgar, D. B., and Trela, J. M. (1976) *J. Bacteriol.* **127**, 1550-1557.
5. Kaledin, A. S., Slyusarenko, A. G., and Gorodetskii, S. I. (1980) *Biokhimiya* **45**, 644-651.
6. Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Keith, A. D., Rigby, P. W. J., Elhai, J., and Hanahan, D. (1988) *Nucleic Acids Res.* **16**, 1563-1575.
7. Murizna, N. V., Vorozheykin, D. P., and Matviienko, N. I. (1988) *Nucleic Acids Res.* **16**: 8172.

*F. C. Lawyer, S. Stoffel, and D. H. Gelfand, unpublished observations.

8. Iijima, S., Uozumi, T., and Beppu, T. (1986) *Agric. Biol. Chem.* **50**, 589-592
9. Nagahari, K., Koshikawa, T., and Sakaguchi, K. (1980) *Gene (Amst.)* **10**, 137-145
10. Slatko, B. E., Benner, J. S., Jager-Quinton, L., Moran, L. S., Simcox, T. G., Vann Cott, E. M., and Wilson, G. G. (1987) *Nucleic Acids Res.* **15**, 9781-9796
11. Kelley, W. S., Chalmers, K., and Murray, N. E. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5632-5636
12. Lin, T. C., Rush, J., Spicer, E. K., and Königsberg, W. H. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7000-7004
13. O'Farrell, P. H., Polisky, B., and Gelfand, D. H. (1978) *J. Bacteriol.* **134**, 645-654
14. Nishiyama, M., Matsubara, N., Yamamoto, K., Iijima, S., Uozumi, T., and Beppu, T. (1986) *J. Biol. Chem.* **261**, 14178-14183
15. Kunai, K., Machida, M., Matsuzawa, H., and Ohta, T. (1986) *Eur. J. Biochem.* **160**, 433-440
16. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., and Oshima, T. (1984) *J. Biol. Chem.* **259**, 2956-2960
17. Brutlag, D., Atkinson, M. B., Setlow, P., and Kornberg, A. (1969) *Biochem. Biophys. Res. Commun.* **37**, 982-989
18. Freemont, P. S., Ollis, D. L., Seitz, T. A., and Joyce, C. M. (1986) *Proteins Struct. Funct. Genet.* **1**, 66-73
19. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature* **313**, 762-766
20. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M., and Steitz, T. A. (1988) *Science* **240**, 199-201
21. Ollis, D. L., Kline, C., and Steitz, T. A. (1985) *Nature* **313**, 818-819
22. Tabor, S., and Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 15330-15333
23. Tabor, S., Huber, H. E., and Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 16212-16223
24. Bernad, A., Zaballos, A., Salas, M., and Blanco, L. (1987) *EMBO J.* **6**, 4219-4225
25. Pizzagalli, A., Valsasini, P., Plevani, P., and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3772-3776
26. Mohan, P. M., Basu, A., Basu, S., Abraham, K. I., and Modak, M. J. (1988) *Biochemistry* **27**, 226-233
27. Basu, A., and Modak, M. J. (1987) *Biochemistry* **26**, 1704-1709
28. Pandey, V. N., Williams, K. R., Stone, K. L., and Modak, M. J. (1987) *Biochemistry* **26**, 7744-7748
29. Pandey, V. N., and Modak, M. J. (1988) *J. Biol. Chem.* **263**, 6068-6073
30. Basu, A., Williams, K. R., and Modak, M. J. (1987) *J. Biol. Chem.* **262**, 9601-9607
31. Joyce, C. M., Ollis, D. L., Rush, J., Steitz, T. A., Königsberg, W. H., and Grindley, N. D. F. (1986) *UCLA Symp. Mol. Cell. Biol. New Ser.* **32**, 197-205
32. Matson, S. W., Capaldo-Kimbrell, F. N., and Bambara, R. A. (1968) *J. Biol. Chem.* **253**, 7851-7856
33. Kelly, W. S., and Grindley, N. D. F. (1976) *Nucleic Acids Res.* **3**, 2971-2984
34. Joyce, C. M., Fujii, D. M., Laks, H. S., Hughes, C. M., and Grindley, N. D. F. (1985) *J. Mol. Biol.* **186**, 283-293
35. Singleton, R., and Amelunxen, R. E. (1973) *Bacteriol. Rev.* **37**, 320-342
36. Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frand, G., and Tratschin, J. D. (1979) *Biochemistry* **18**, 5698-5703
37. Ponnuswamy, P. K., Muthusamy, R., and Manavalan, P. (1982) *Int. J. Biol. Macromol.* **4**, 189-190
38. Young, R. A., and Davis, R. W. (1983) *Science* **222**, 778-782
39. Wood, W. B. (1966) *J. Mol. Biol.* **16**, 118-133
40. Casadaban, M. J., and Cohen, S. N. (1980) *J. Mol. Biol.* **138**, 179-207
41. Cole, G. E., McCabe, P. C., Inlow, D., Gelfand, D. H., Ben-Bassat, A., and Innis, M. A. (1988) *BioTechnology* **6**, 417-421
42. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 201-205, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
43. Marinus, M. G., Caraway, M., Frey, A. Z., Brown, L., and Arraj, J. A. (1983) *Mol. Gen. Genet.* **192**, 288-289
44. Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1194-1198
45. Loenen, W. A. M., and Blattner, F. R. (1983) *Gene (Amst.)* **26**, 171-179
46. Russel, M., Kidd, S., and Kelley, M. R. (1986) *Gene (Amst.)* **45**, 333-338
47. Greene, P. J., Betlach, M. C., Goodman, H. M., and Boyer, H. (1974) *Methods Mol. Biol.* **7**, 87-111
48. Modrich, P., and Zabel, D. (1976) *J. Biol. Chem.* **251**, 5866-5874
49. Bingham, A. H. A., Sharman, A. F., and Atkinson, T. (1977) *FEBS Lett.* **76**, 250-256
50. Bickle, T. A., Pirrotta, V., and Imber, R. (1977) *Nucleic Acids Res.* **4**, 2561-2572
51. Gingras, T. R., Myers, P. A., Olson, J. A., Hanberg, F. A., and Roberts, R. J. (1978) *J. Mol. Biol.* **118**, 113-122
52. Moore, S. K., and James, E. (1976) *Anal. Biochem.* **75**, 545-554
53. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580
54. Birnboim, H. C., and Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523
55. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
56. Castenholz, R. W. (1969) *Bacteriol. Rev.* **33**, 476-504
57. Young, R. A., and Davis, R. W. (1985) *Genetic Engineering* (Setlow, J. K., and Hollaender, A., eds) Vol. 7, pp. 29-41, Plenum Publishing Corp., New York
58. Sheldon, E. L., Kellogg, D. E., Watson, R., Levenson, C. H., and Erlich, H. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 9085-9089
59. Huynh, T. V., Young, R. A., and Davis, R. W. (1985) in *DNA Cloning, IRL Press, Oxford Techniques: A Practical Approach* (Glover, D., ed) pp. 50-77, IRL Press, Oxford
60. Wong, G., Arnhem, N., Clark, R., McCabe, P., Innis, M., Aldwin, L., Nitecki, D., and McCormick, F. (1986) *Cancer Res.* **46**, 6029-6033
61. Mack, D. H., Bloch, W., Nath, N., and Sninsky, J. J. (1988) *J. Virol.* **62**, 4786-4790
62. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6**, 2932-2943
63. Goldstein, L. S. B., Laymon, R. A., and McIntosh, J. R. (1986) *J. Cell Biol.* **102**, 2076-2087
64. Messing, J. (1983) *Methods Enzymol.* **101**, 20-78
65. Sanger, F. (1981) *Science* **214**, 1205-1210
66. Barnes, W. M., Bevan, M., and Son, P. H. (1983) *Methods Enzymol.* **101**, 98-122
67. Gough, J. A., and Murray, N. E. (1983) *J. Mol. Biol.* **166**, 1-19
68. Mizusawa, S., Nishimura, S., and Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319-1324
69. Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. A. D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9436-9440
70. Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951-2961

Continued on next page.

Table V
DNA Sequencing Primers

Primer	Taq Pol I Nucleotide Coordinates ^a
MK12	625 → 637
MK13	629 → 634
MK14	630 → 639
MK15	630 → 637
MK16	2360 → 2367
MK17	637 → 645
MK18	625 → 631
MK19	1740 → 1749
MK20	1321 → 1609
MK21	291 → 309
MK22	291 → 311
MK23	353 → 373
MK24	1013 → 1012
MK25	136 → 156
MK26	24 → 42
MK27	601 → 605
MK28	1313 → 1315
MK29	1313 → 1315
MK30	2340 → 2349
MK31	2340 → 2359
MK32	351 → 357
MK33	2361 → 2360
MK34	597 → 598
MK35	130 → 148
MK36	64 → 48

^aArrows denote strand sequenced with each primer: → is sense strand, 5' to 3', and ← is antisense strand, 3' to 5'.

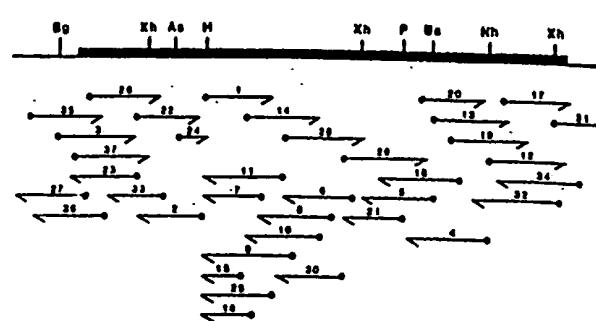


Fig. 8. DNA sequencing strategy. The bold line on the map delineates the coding sequencing for *Taq Pol I*. Arrows indicate sequence obtained in the sense (+) or anti-sense (-) direction. Length of the arrows corresponds to the amount of sequence obtained in each case. 100% of the DNA sequence was determined on both strands. 1-3, the universal sequencing primer was used for sequencing on templates pFCN23S, pFCN3, and the B fragment BglII-HindIII subclone. 4-11, the universal sequencing primer was used with *lacZ* deletion templates AABm #1, AABm #13, AABm #30, #32, #33, #34, and #39. 12-13, reverse sequencing primer was used with A-fragment deletions Afp-20-SmaI and Afp-20-BamH. 14-37, primers utilized were MK12-124, MK130-136, MK138-143, MK145, MK150, MK151, MK153, MK154, and MK159.